

Conference Abstract

Metadata Standards for Genomic Sequence Data: Past and Future of MlxS Standards Family

Pelin Yilmaz ‡

‡ Max Planck Institute for Marine Microbiology, Bremen, Germany

Corresponding author: Pelin Yilmaz (pyilmaz@mpi-bremen.de)

Received: 18 Aug 2017 | Published: 18 Aug 2017

Citation: Yilmaz P (2017) Metadata Standards for Genomic Sequence Data: Past and Future of MlxS Standards Family. Proceedings of TDWG 1: e20423. <https://doi.org/10.3897/tdwgproceedings.1.20423>

Abstract

Data-intensive microbial (ecology) research requires approaches to organize information in logical formats to enable comparative studies, replication of experiments, and development of informatics tools. As the quantity and diversity of genomic data continues to increase at an exponential rate, it is imperative that these data are findable, accessible, interoperable, and reusable through the use of a standard format. Since 2005, the Genomics Standards Consortium (GSC) operates an international open-membership working body of over 500 researchers from 15 countries promoting community-driven efforts for the reuse and analysis of contextual metadata describing the collected sample, the environment and/or the host, and sequencing methodologies and technologies. GSC's development and implementation of Minimum Information Standards (MlxS) established a community-based mechanism for sharing genomic data through a common framework. The GSC's suite of genomic standards (Field et al. 2008, Yilmaz et al. 2011) have been supported for over a decade by the International Nucleotide Sequence Database Collaboration (INSDC) databases, thus allowing for a complete environmental and experimental description of sequenced samples (Barrett et al. 2012, Cochrane et al. 2015). While broadly accepted and used by the microbial (ecology) research community, MlxS has several shortcomings, as well as areas that require further development. The GSC is committed to engaging domain experts, in order to: (i) expand coverage and breadth to accommodate new data types and emerging technologies, (ii) maximize usability, (iii) expedite further evolution according to community needs, and (iv) automatize update of MlxS.

Keywords

omics data, microbial ecology, INSDC, checklist, standards

Presenting author

Pelin Yilmaz

Max Planck Institute for Marine Microbiology, Microbial Genomics and Bioinformatics Research Group

References

- Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A (2012) NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Research 41: D991-D995. <https://doi.org/10.1093/nar/gks1193>
- Cochrane G, Karsch-Mizrachi I, Takagi T, Database Collaboration INS (2015) The International Nucleotide Sequence Database Collaboration. Nucleic Acids Research 44: D48-D50. <https://doi.org/10.1093/nar/gkv1323>
- Field D, Garrity G, Sansone S, Sterk P, Gray T, Kyrpides N, Hirschman L, Glöckner FO, Kottmann R, Angiuoli S, White O, Dawyndt P, Thomson N, Gil IS, Morrison N, Tatusova T, Mizrachi I, Vaughan R, Cochrane G, Kagan L, Murphy S, Schriml L (2008) Meeting report: the Fifth Genomic Standards Consortium (GSC) Workshop. OMICS: A Journal of Integrative Biology 12 (2): 109-113. <https://doi.org/10.1089/omi.2008.a3b3>
- Yilmaz P, Gilbert JA, Knight R, Amaral-Zettler L, Karsch-Mizrachi I, Cochrane G, Nakamura Y, Sansone S, Glöckner FO, Field D (2011) The genomic standards consortium: bringing standards to life for microbial ecology. The ISME Journal 5 (10): 1565-1567. <https://doi.org/10.1038/ismej.2011.39>